London School of Economics and Political Science
Statistical Models and Data Analysis
ST201/ST203.1 Colin Chalmers

Computer Workshop 8

In this workshop we shall:

- *use SPSS plotting routines to investigate time trends in data*
- *review regression for estimating relationships*
- *see how SPSS handles Time Series*

Computer workshop 7 showed how a fairly crude estimation procedure in Excel could decompose a Time Series into trend and seasonal effects. The work involved however was considerable which suggests, as indicated at the start of this course, that as the work involved becomes more complex it is worthwhile switching to a custom-built package rather than a general portmanteau program. It is possible to build macros to deal with complex operations but 'why re-invent the wheel'.

The case study which Malcolm De Silva was concerned with - the Brentstone Horticultural Show will be used again.

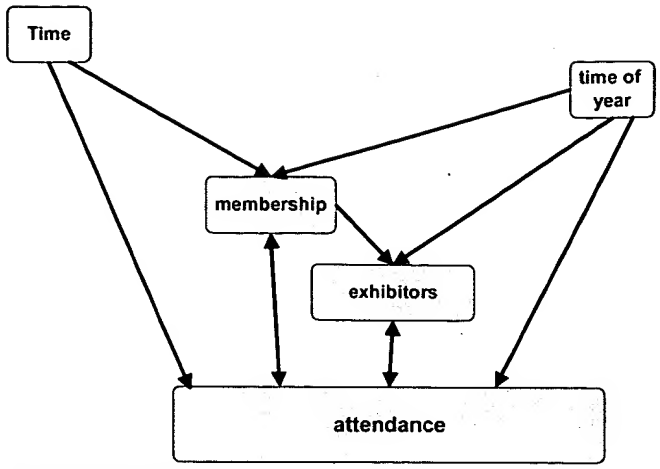
Open workshop8.sav and look at the data in both its views.

The variables in this file are self-explanatory. *Years*, a coded form for years, *shows* coded 1..3 (turn on the value labels to see what 1..3 mean), *time* since the beginning of 1994 in months, a special short text variable which will help in plotting, three dummy flags for the shows and three measured, inter-dependent outcomes variables, *number of members*, *number of exhibitors* and *number of people* attending each show.

Malcolm has already looked at *exhibitors* and *attendance* and has produced forecasts. He will look again at these aspects and also try to model the whole process. The process, he believes, runs as follows.

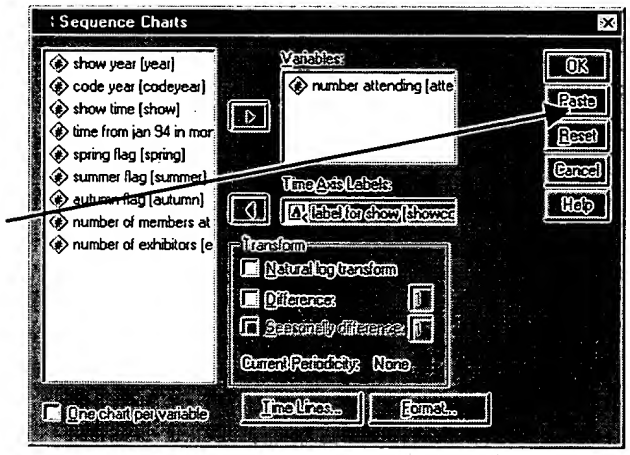
Over time there has been an increasing interest in horticulture with the advent of gardening and makeover programs. This will lead to an increase in membership of societies such at Brentstone and attendance at local shows. This growth will be time dependent though in a seasonal way as certain times of the year will give a greater appeal. As membership increases so too will the number of exhibitors and both will come to the shows so they will influence final attendance. It is however a moot point as to whether this effect will simply follow from the increasing interest over time. Since all three of these activities give income to the society, forecasting any one will be beneficial whilst the influence of controllable factors such as membership (which can at least be encouraged through local advertising) and exhibitors will allow cost benefit analysis. Finally, though this will not at this stage be considered there will be a flow back into the system in that those attending the show will be more likely to become members and exhibitors in the future.

To summarise this opinion, Malcolm produces an influence diagram which any full-blooded analysis would investigate.



1. time series charting

SPSS has a time-series charting facility. *<graph><sequence>* Enter the *number attending* as the outcome and *label for show* as the Time Axis label.
Do not run this but ***Paste*** it into a new syntax file.



Insert comments (*/ .) in this file to indicate who wrote it and what each set of instructions does. Save it to a named file. Hand in a hard copy of your syntax file with the comments asked (questions) embedded in it within two days.

Run the routine *(recall, go into the syntax file, set the cursor in a block of instruction or block mark a whole set and press the run arrow)*

Qu.1: Is this plot as easy to perform as the equivalent in Excel and can you see how it can be edited?

Next copy the block of text in the syntax file which carries out this plot to a position below (watch for that '.') ,edit out *attend* replacing it with the two variables *members* and *exhibito* and rerun.(**Hint**: the pasting is done easily by using the utilities menu. This lists the variables and the names can be pasted directly into the syntax file to save mistyping.)

Qu.2: Is there a relationship between membership and exhibitors?

Confirm this with a scatterplot of *members* versus *exhibito* ***(remember to paste this into your syntax file as a record!)***

2.Modelling the outcomes

Malcolm can now see that there is some relationship between the three outcome variables, that there is some seasonal variation and certainly an upward trend with time. So he fits some regression models.

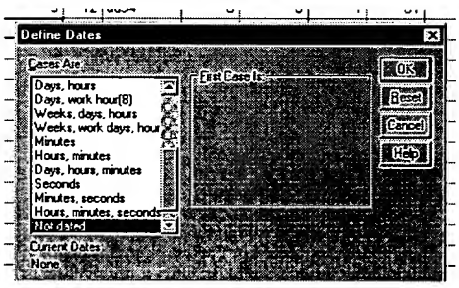
Fit a stepwise regression of *attend* on the two drivers *members* and *exhibito* and the time variable , *time*. (remember to update your syntax file)
Qu.3: How much influence do *members* and *exhibito* having on *attend* once the general time trend is taken into account? Does this imply that financially it is not worthwhile encouraging a growth in membership above the natural trend?

Next Malcolm looks at the effect of season using a general linear model approach. He builds a main effects model for *attend* using again *members*, *exhibito* and *time* as covariables but also including *show* as a factor. He then reduces this model by removing gradually (i.e. one at a time) any covariable or factor which is not statistically significant and re-running. Follow his example.

Qu.4: What is the final model, how good is it and what are its implications? (Hint: you may find setting some of the *options* helpful - in particular descriptive statistics and parameter estimates. Also, once you have pasted into the syntax it is very easy to change the model.)

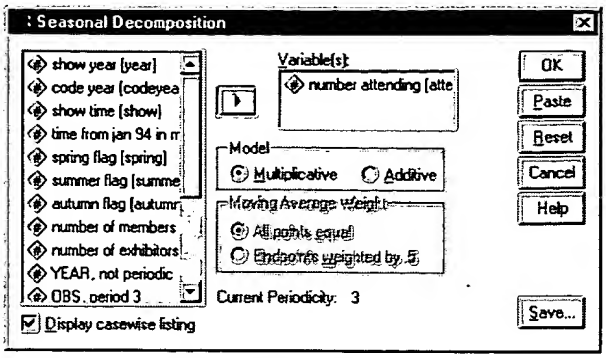
3. Using Seasonal Decomposition Time Series Programs

The trend and seasonal model proved successful in Excel. SPSS has a slightly different approach in that it does not assume a linear trend but estimates the seasonal factors using simply a moving average trend and then the user can produce both the trend and the seasons to obtain a forecast. **But first a snag!**
Try *<analyze><time series><seasonal decomposition>* **bother!!!**



Although we have a number of obvious time dependent variables such as *year*, *show* etc SPSS insists that we use an inbuilt series of its own choosing. You might be able to define dates suitable for SPSS in the data window. But...
When you try to do this the only options are for bog-standard seasons such as years, months, weeks and days - no mention of four-monthly intervals. There are a number of situations in which SPSS has facilities but they cannot be accessed from the menu screen - hence the insistence that syntax files should be used. So go back into your syntax file, insert and run the following line.
DATE YEAR 1994 OBServation 1 3.
This tells SPSS that the data setup is a time series with increasing years from 1994 and with three observations labelled 1,2,3 within each year.
Next look at the data file and see the new system variables that SPSS has created.

Now try *<Analyze><Time Series><Seasonal Decomposition>* again. You should find that a screen does open and indicates that you have successfully created a season of three periods per year. Note that I have checked the *Display casewise listing* box.

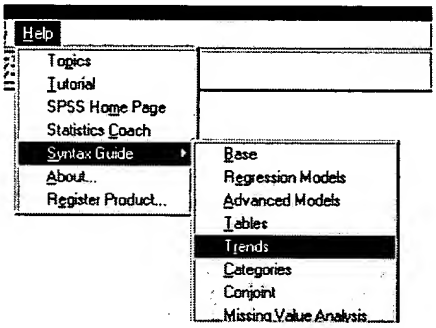


Paste this into the syntax file and run all the commands following the words *seasonal decomposition* i.e. block mark all the instructions before pressing the run arrow.

```
* Seasonal Decomposition.
TSET PRINT=DEFAULT NEWVAR=ALL .
SEASON
 /VARIABLES=attend
 /MODEL=MULTIPLICATIVE.
```

Look at the outputs and you will see that it starts with a centred moving average, follows this with estimated yearly seasons, converts these to equal seasons per year etc.

Now refer to the help menu on TRENDS to interpret the syntax and see how it could be altered.



Qu.5: How would you change from a multiplicative model to an additive one and how would the output differ? (try it if you like)

Qu.6: What changes would be necessary if the data had been quarterly rather than every four months?

To forecast, it is simplest to take the detrended series and the seasonal terms back into Excel and then use any appropriate trend to extend the series and multiply(or add) the appropriate seasonal terms.

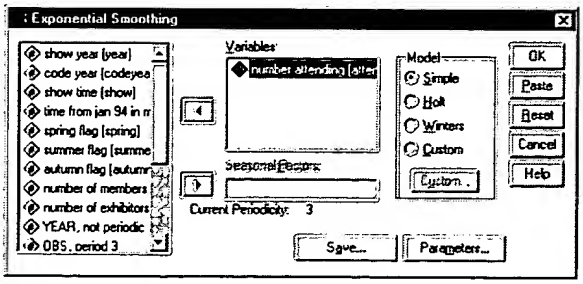
Look at the data now and see what has been added.

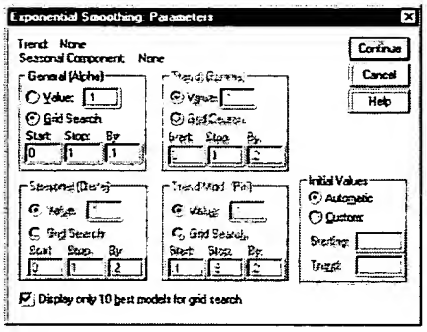
Qu.7: What do the variables err_1, sas_1, saf_1 and sft_1 represent (compare with the output file)?

4. Other approaches.

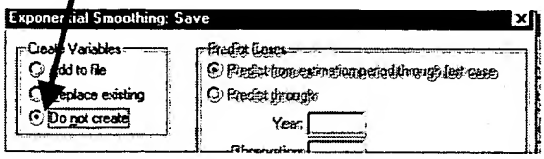
Now try the other option under Time Series which we have met already, *Exponential smoothing*. We experimented with this in Excel. SPSS allows us the same pleasure.

Open the window and you will find that SPSS remembers that we have a four-monthly seasonal figure and a time variable.





Since you don't know where to start the series or what smoothing parameter to use (exactly as in Excel) let SPSS try ten values for the Alpha and select its own starting value. This can be done by using the parameters button and setting the flag for a grid search. Meanwhile in the save screen specify that new variables are not to be created or forecasts produced.



Paste this syntax and run it. Decide on an optimum alpha and rerun but forecast ahead. Rather than meddle with the menu screen, edit the syntax file inserting the value for alpha which you think most suitable which I have marked with XXX

```
* Exponential Smoothing.
TSET PRINT=DEFAULT NEWVAR=ALL .
use year 1994 observation 1 thru year 2001 observation 1.
PREDICT THRU YEAR 2001 OBSERVATION 2 .
EXSMOOTH /VARIABLES=attend
 /MODEL=NN
 /ALPHA= XXX
 /INITIAL=CALCULATE.
```

note this

your value replaces

Qu.8 What is your forecast for summer 2001? What would it be for 2002? (Either think about it or try changing the syntax).

The Holt and Winter models as given in the menu allow further forms of smoothing to reflect trend and seasonality rather than the simple one-step ahead procedure of exponential smoothing. Extra parameters, (Gamma for the Holt model, Gamma and Delta for the Holt-Winters model) are introduced to give flexibility but at the expense of increased complexity. We try just one to finish.

If you have looked at the *help* file on TRENDS you will see that the parameter for MODEL, NN, can be replaced by a combination of N,L,E and D standing for No, Linear , Exponential and Damped trend with N, A or M standing for No, Additive or Multiplicative season. Try just one example as shown. Then add the necessary extra values to *showcode* and repeat a sequence plot with attend and the latest fitted variable.

```
TSET PRINT=DEFAULT NEWVAR=all .
use year 1994 observation 1 thru year 2001
observation 1.
PREDICT THRU YEAR 2002 OBSERVATION 3 .
EXSMOOTH /VARIABLES=attend
 /SEASFACT=obs_
 /MODEL=LM
 /ALPHA=0.5
 /DELTA= 1.0
 /GAMMA= 0.75
 /INITIAL=CALCULATE.
```

Qu.9: What are your forecasts now for the coming year?

Qu.10: Why is the forecast routine having particular difficulty? (look at how it lags behind the actual values)